

Anytime Reliable Codes for Stabilizing Plants over Erasure Channels

Ravi Teja Sukhavasi

Babak Hassibi

Abstract—The problem of stabilizing an unstable plant over a noisy communication link is an increasingly important one that arises in problems of distributed control and networked control systems. Although the work of Schulman and Sahai over the past two decades, and their development of the notions of “tree codes” and “anytime capacity”, provides the theoretical framework for studying such problems, there has been scant practical progress in this area because explicit constructions of tree codes with efficient encoding and decoding did not exist. To stabilize an unstable plant driven by bounded noise over a noisy channel one needs real-time encoding and real-time decoding and a reliability which increases exponentially with delay, which is what tree codes guarantee. We prove the existence of linear tree codes with high probability and, for erasure channels, give an explicit construction with an expected encoding and decoding complexity that is constant per time instant. We give sufficient conditions on the rate and reliability required of the tree codes to stabilize vector plants and argue that they are asymptotically tight. This work takes a major step towards controlling plants over noisy channels, and we demonstrate the efficacy of the method through several examples.

I. INTRODUCTION

In control theory, the output of a dynamical system is observed and a controller is designed to regulate its behavior. The controller needs to react and generate control signals in real-time. In most traditional control systems, the controller and the plant are colocated and hence there is no measurement loss. There are increasingly many applications such as networked control systems [1] and distributed computing [2] where systems are remotely controlled and where measurement and control signals are transmitted across noisy channels. This necessitates a need to *reliably* communicate the measurement and control signals by correcting for the errors introduced by the channels. Although Shannon’s information theory is concerned with reliable transmission of a message from one point to another over a noisy channel, the reliability is achieved at the price of large delays which may lead to instability when they occur in the feedback loop of a control system. Hence, we need practical real-time encoding and decoding schemes with appropriate reliability for controlling systems over lossy networks.

Consider a control system with a single observer that communicates with the controller over a lossy communication channel and where the feedback link from the controller to the plant is noiseless. When the channel is rate-limited and deterministic, significant progress has been made (see eg., [3], [4]) in understanding the bandwidth requirements for stabilizing open loop unstable systems. When the communication channel is stochastic, [5] provides a necessary and sufficient condition on the communication reliability needed over such a channel

to stabilize an unstable scalar linear process, and proposes the notion of feedback anytime capacity as the appropriate figure of merit for such channels. In essence, the encoder is causal and the probability of error in decoding a source symbol that was transmitted d time instants ago should decay exponentially in the decoding delay d .

Although the connection between communication reliability and control is clear, very little is known about error-correcting codes that can achieve such reliabilities. Prior to the work of [5], and in a different context, [2] proved the existence of codes which under maximum likelihood decoding achieve such reliabilities and referred to them as tree codes. Note that any real-time error correcting code is causal and since it encodes the entire trajectory of a process, it has a natural tree structure to it. [2] proves the existence of nonlinear tree codes yet gives no explicit constructions and/or efficient decoding algorithms. Much more recently [6] proposed efficient error correcting codes for unstable systems where the state grows only polynomially large with time. So, for linear unstable systems that have an exponential growth rate, all that is known in the way of error correction is the existence of tree codes which are, in general, non-linear. Moreover, the existence results are not with a “high probability”. When the state of an unstable scalar linear process is available at the encoder, [7] and [8] develop encoding-decoding schemes that can stabilize such a process over the binary symmetric channel and the binary erasure channel respectively. But little is known in the way of stabilizing partially observed vector-valued processes over a stochastic communication channel.

The subject of error correcting codes for control is in its relative infancy, much as the subject of block coding was after Shannon’s seminal work in [9]. So, a first step towards realizing practical encoder-decoder pairs with anytime reliabilities is to explore linear encoding schemes. We consider rate $R = \frac{k}{n}$ causal linear codes which map a sequence of k -dimensional binary vectors $\{b_\tau\}_{\tau=0}^\infty$ to a sequence of n -dimensional binary vectors $\{c_\tau\}_{\tau=0}^\infty$ where c_τ is only a function of $\{b_\tau\}_{\tau=0}^t$. Such a code is anytime reliable if there exist constants $\beta > 0, \eta > 0$ and a delay $d_o > 0$ such that at all times t , $P(\hat{b}_{t-d}|t \neq b_{t-d}) \leq \eta 2^{-\beta nd}$.

The contributions of this paper are as follows: 1. We show that linear tree codes exist and further, that they exist with a high probability. 2. For the binary erasure channel, we propose a maximum likelihood decoder whose average complexity of decoding is constant per each time iteration and for which the probability that the complexity at a given time t exceeds KC^3 decays exponentially in C . 3. We also prove asymptotically tight

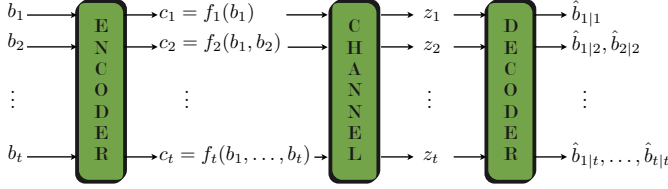


Fig. 1. Causal encoding and decoding

sufficient conditions on the rate R and exponent β needed to stabilize vector-valued processes over a noisy channel. As a consequence, we can efficiently stabilize a partially observed unstable linear process over a binary erasure channel without any channel feedback.

In Section II, we introduce the notation and set up the problem. In Section III, we introduce the ensemble of time invariant codes and show that they are anytime reliable with a high probability. In Section IV, we present a simple decoding algorithm for the BEC and in Section V, we derive sufficient conditions for stabilizing unstable linear systems over noisy channels. We present some simulations in Section VIII to demonstrate the efficacy of the decoding algorithm.

II. PROBLEM SETUP

We will begin by introducing some notation

- 1) For any matrix F , $\bar{F} \triangleq \text{abs}(F)$, i.e., $\bar{F}_{i,j} = |F_{i,j}|, \forall i, j$
- 2) $\lambda(F) \triangleq$ largest eigen value of F in magnitude.
- 3) For a vector x , $x^{(i)}$ denotes the i^{th} component of x .
- 4) $\mathbf{1}_m \triangleq [1, \dots, 1]^T$, i.e., a column with m 1's.
- 5) For $w, v \in \mathbb{R}^m$, $w \geq v$ denotes component-wise inequality.

Consider the following m -dimensional unstable linear system with scalar measurements. Assuming that the system is observable, without loss of generality, it can be cast in the following canonical form.

$$x_{t+1} = Fx_t + Bu_t + w_t, \quad y_t = Hx_t + v_t \quad (1)$$

where

$$F = \begin{bmatrix} -a_1 & 1 & 0 & \dots \\ -a_2 & 0 & 1 & 0 \\ \vdots & \vdots & & \ddots \\ -a_{m-1} & \dots & \dots & 0 & 1 \\ -a_m & 0 & \dots & \dots & 0 \end{bmatrix}, \quad H = [1, 0, \dots, 0]$$

where $\lambda(F) > 1$, u_t is the control input and, w_t and v_t are bounded process and measurement noise variables, i.e., $\|w_t\|_\infty < \frac{W}{2}$ and $\|v_t\|_\infty < \frac{V}{2}$. Note that the characteristic polynomial of F is $z^n + a_1 z^{n-1} + \dots + a_m$.

The measurements $\{y_t\}$ are made by an observer while the control inputs $\{u_t\}$ are applied by a remote controller that is connected to the observer by a noisy communication channel. Naturally, the measurements $y_{0:t-1}$ will need to be encoded by the observer to provide protection from the noisy channel while the controller will need to decode the channel outputs to estimate the state x_t and apply a suitable control input u_t . This can be accomplished by employing a channel encoder at the observer and a decoder at the controller. For simplicity, we

will assume that the channel input alphabet is binary. Suppose one time step of system evolution in (1) corresponds to n channel uses¹. Then, at each instant of time t , the operations performed by the observer, the channel encoder, the channel decoder and the controller can be described as follows. The observer generates a k -bit message, $b_t \in \{0, 1\}^k$, that is a causal function of the measurements, i.e., it depends only on $y_{0:t}$. Then the channel encoder causally encodes $b_{0:t} \in \{0, 1\}^{kt}$ to generate the n channel inputs $c_t \in \{0, 1\}^n$. Note that the rate of the channel encoder is $R = k/n$. Denote the n channel outputs corresponding to c_t by $z_t \in \mathcal{Z}^n$, where \mathcal{Z} denotes the channel output alphabet. Using the channel outputs received so far, i.e., $z_{0:t} \in \mathcal{Z}^{nt}$, the channel decoder generates estimates $\{\hat{b}_{\tau|t}\}_{\tau \leq t}$ of $\{b_\tau\}_{\tau \leq t}$, which, in turn, the controller uses to generate the control input u_{t+1} . This is illustrated in Fig. 1. *Note that we do not assume any channel feedback.* Now, define

$$P_{t,d}^e = P\left(\min\{\tau : \hat{b}_{\tau|t} \neq b_\tau\} = t - d + 1\right)$$

Thus, $P_{t,d}^e$ is the probability that the earliest error is d steps in the past.

Definition 1 (Anytime reliability): We say that an encoder-decoder pair is (R, β, d_o) -anytime reliable if

$$P_{t,d}^e \leq \eta 2^{-n\beta d}, \quad \forall t, d \geq d_o \quad (2)$$

In some cases, we write that a code is (R, β) -anytime reliable. This means that there exists a fixed $d_o > 0$ such that the code is (R, β, d_o) -anytime reliable.

We will show in Section V (Theorem 5.1) that (R, β) -anytime reliability is a sufficient condition to stabilize (1) in the mean squared sense². In what follows, we will demonstrate causal linear codes which under maximum likelihood decoding achieve such exponential reliabilities.

III. LINEAR ANYTIME CODES

As discussed earlier, a first step towards developing practical encoding and decoding schemes for automatic control is to study the existence of linear codes with anytime reliability. We will begin by defining a causal linear code.

Definition 2 (Causal Linear Code): A causal linear code is a sequence of linear maps $f_\tau : \{0, 1\}^{k\tau} \mapsto \{0, 1\}^n$, $\tau \geq 0$ and hence can be represented as

$$f_\tau(b_{1:\tau}) = G_{\tau 1} b_1 + G_{\tau 2} b_2 + \dots + G_{\tau \tau} b_\tau \quad (3)$$

where $G_{ij} \in \{0, 1\}^{n \times k}$

We denote $c_\tau \triangleq f_\tau(b_{1:\tau})$. Note that a tree code is a more general construction where f_τ need not be linear. Also note that the associated code rate is $R = \frac{k}{n}$. The above encoding is equivalent to using a semi-infinite dimensional block lower triangular generator matrix, $G_{n,R}$, whose entries are clear from (3) or equivalently as a semi-infinite dimensional block lower triangular parity check matrix, $\mathbb{H}_{n,R}$ (the parity check matrix

¹In practice, the system evolution in (1) is obtained by discretizing a continuous time differential equation. So, the interval of discretization could be adjusted to correspond to an integer number of channel uses, provided the channel use instances are close enough.

²can be easily extended to any other norm

satisfies $H_{n,R}G_{n,R} = 0$.)

$$\mathbb{H}_{n,R} = \begin{bmatrix} H_{11} & 0 & \dots & \dots & \dots \\ H_{21} & H_{22} & 0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_{\tau 1} & H_{\tau 2} & \dots & H_{\tau \tau} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (4)$$

where³ $H_{ij} \in \{0, 1\}^{\bar{n} \times n}$ and $\bar{n} = n(1 - R)$. In order to ensure that the code rate is equal to the design rate $R = \frac{k}{n}$, $\mathbb{H}_{n,R}^t$ needs to be full rank for every t , where $\mathbb{H}_{n,R}^t$ is the $\bar{n}t \times nt$ leading principal minor of $\mathbb{H}_{n,R}$. This will happen if H_{ii} is full rank for all i . The existence results that follow imply the existence of anytime reliable $\mathbb{H}_{n,R}$ whose code rate is same as the design rate.

We will present all our results for binary input output symmetric channels⁴. The Bhattacharya parameter ζ for such channels is defined as

$$\zeta = \int_{-\infty}^{\infty} \sqrt{p(z|X=1)p(z|X=0)} dz$$

where z and X denote the channel output and input, respectively. We will begin by proving the existence of such codes that are (R, β) -anytime reliable over a finite time horizon, T , i.e., under ML decoding $P_{d,t}^e \leq \eta 2^{-\beta d}$, $\forall t \leq T$. We will then prove their existence for all time. Due to space limitations, proofs for all the results in this section are presented in a companion paper, [10].

A. Finite Time Horizon

Over a finite time horizon, T , a causal linear code is represented by a block lower triangular parity check matrix $\mathbb{H}_{n,R,T} \in \{0, 1\}^{\bar{n}T \times nT}$. The following Theorem guarantees the existence of a $\mathbb{H}_{n,R,T}$ that is (R, β) -anytime reliable.

Theorem 3.1: For each time $T > 0$, rate R and exponent β such that

$$R < 1 - \log_2(1 + \zeta), \quad \text{and} \\ \beta < H^{-1}(1 - R) \left(\log_2 \left(\frac{1}{\zeta} \right) + \log_2 [2^{1-R} - 1] \right)$$

there exists a causal linear code $H(n, k, T)$ that is (R, β) -anytime reliable.

$H^{-1}(1 - R)$ is the smaller root of the equation $H(x) = 1 - R$, where $H(\cdot)$ is the binary entropy function. Theorem 3.1 proves the existence of finite dimensional causal linear codes, $\mathbb{H}_{n,R,T}$, that are anytime reliable for decoding instants upto time T . In the following subsection, we demonstrate the existence of semi-infinite causal linear codes, $\mathbb{H}_{n,R}$, that are anytime reliable for all decoding instants. We also show that such codes drawn from an appropriate ensemble are anytime reliable with a high probability. The key is to impose a Toeplitz structure on the parity check matrix.

³While for a given generator matrix, the parity check matrix is not unique, when $\mathbb{G}_{n,R}$ is block lower, it is easy to see that $\mathbb{H}_{n,R}$ can also be chosen to be block lower.

⁴which can be easily extended to more general memoryless channels

B. Time Invariant Codes

Consider causal linear codes with the following Toeplitz structure

$$\mathbb{H}_{n,R}^{TZ} = \begin{bmatrix} H_1 & 0 & \dots & \dots & \dots \\ H_2 & H_1 & 0 & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H_\tau & H_{\tau-1} & \dots & H_1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The superscript TZ in $\mathbb{H}_{n,R}^{TZ}$ denotes ‘Toeplitz’. $\mathbb{H}_{n,R}^{TZ}$ is obtained from $\mathbb{H}_{n,R}$ in (4) by setting $H_{ij} = H_{i-j+1}$ for $i \geq j$. Due to the Toeplitz structure, we have the following invariance, $P_{t,d}^e = P_{t',d}^e$ for all t, t' . The notion of time invariance is analogous to the convolutional structure used to show the existence of infinite tree codes in [2]. The code $\mathbb{H}_{n,R}^{TZ}$ will be referred to as a time-invariant code. This time invariance obviates the need to union bound over all time t and hence allows us to prove that such codes which are anytime reliable are abundant.

Definition 3 (The ensemble \mathbb{TZ}_p): The ensemble \mathbb{TZ}_p of time-invariant codes, $\mathbb{H}_{n,R}^{TZ}$, is obtained as follows, H_1 is any full rank binary matrix and for $\tau \geq 2$, the entries of H_τ are chosen i.i.d according to Bernoulli(p), i.e., each entry is 1 with probability p and 0 otherwise.

Note that H_1 being full rank implies that $H_{n,R}^t$ is full rank for every t . For the ensemble \mathbb{TZ}_p , we have the following result

Theorem 3.2 (Abundance of time-invariant codes): For any rate R and exponent β such that

$$R < 1 - \frac{\log_2(1 + \zeta)}{\log_2(1/(1-p))}, \quad \text{and} \\ \beta < H^{-1}(1 - R) \left(\log_2 \left(\frac{1}{\zeta} \right) + \log_2 [(1-p)^{-(1-R)} - 1] \right)$$

if $\mathbb{H}_{n,R}^{TZ}$ is chosen from \mathbb{TZ}_p , then

$$P(\mathbb{H}_{n,R}^{TZ} \text{ is } (R, \beta, d_o) \text{ - anytime reliable}) \geq 1 - 2^{-\Omega(nd_o)}$$

Note that by choosing p small, we can trade off better rates and exponents with sparser parity check matrices. Note that for BEC(ϵ), $\zeta = \epsilon$ and for BSC(ϵ), $\zeta = 2\sqrt{\epsilon(1-\epsilon)}$. For the Binary Symmetric Channel (BSC) with bit flip probability ϵ and for $p = \frac{1}{2}$, the threshold for rate in Theorem 3.2 becomes $R < 1 - 2\log_2(\sqrt{\epsilon} + \sqrt{1-\epsilon})$. It turns out that this can be strengthened as follows.

Theorem 3.3 (Tighter bounds for BSC(ϵ)): For any rate R and exponent β such that

$$R < 1 - H(\epsilon), \quad \beta < KL(H^{-1}(1 - R) \parallel \min\{\epsilon, 1 - \epsilon\})$$

if $\mathbb{H}_{n,R}^{TZ}$ is chosen from $\mathbb{TZ}_{\frac{1}{2}}$, then

$$P(\mathbb{H}_{n,R}^{TZ} \text{ is } (R, \beta, d_o) \text{ - anytime reliable}) \geq 1 - 2^{-\Omega(nd_o)}$$

IV. DECODING OVER THE BEC

Owing to the simplicity of the erasure channel, it is possible to come up with an efficient way to perform maximum likelihood decoding at each time step. We will show that the average

complexity of the decoding operation at any time t is constant and that it being larger than KC^3 decays exponentially in C . Consider an arbitrary decoding instant t , let $c = [c_1^T, \dots, c_t^T]^T$ be the transmitted codeword and let $z = [z_1^T, \dots, z_t^T]^T$ denote the corresponding channel outputs. Recall that $\mathbb{H}_{n,R}^t$ denotes the $\bar{n}t \times nt$ leading principal minor of $\mathbb{H}_{n,R}$. Let z_e denote the erasures in z and let H_e denote the columns of $\mathbb{H}_{n,R}^t$ that correspond to the positions of the erasures. Also, let \tilde{z}_e denote the unerased entries of z and let \tilde{H}_e denote the columns of $\mathbb{H}_{n,R}^t$ excluding H_e . So, we have the following parity check condition on z_e , $H_e z_e = \tilde{H}_e \tilde{z}_e$. Since \tilde{z}_e is known at the decoder, $s \triangleq \tilde{H}_e \tilde{z}_e$ is known. Maximum likelihood decoding boils down to solving the linear equation $H_e z_e = s$. Due to the lower triangular nature of H_e , unlike in the case of traditional block coding, this equation will typically not have a unique solution, since H_e will typically not be full rank. This is alright as we are not interested in decoding the entire z_e correctly, we only care about decoding the earlier entries accurately. If $z_e = [z_{e,1}^T, z_{e,2}^T]^T$, then $z_{e,1}$ corresponds to the earlier time instants while $z_{e,2}$ corresponds to the latter time instants. The desired reliability requires one to recover $z_{e,1}$ with an exponentially smaller error probability than $z_{e,2}$. Since H_e is lower triangular, we can write $H_e z_e = s$ as

$$\begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix} \begin{bmatrix} z_{e,1} \\ z_{e,2} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (6)$$

Let $H_{e,22}^\perp$ denote the orthogonal complement of $H_{e,22}$, i.e., $H_{e,22}^\perp H_{e,22} = 0$. Then multiplying both sides of (6) with $\text{diag}(I, H_{e,22}^\perp)$, we get

$$\begin{bmatrix} H_{e,11} \\ H_{e,22}^\perp H_{e,21} \end{bmatrix} z_{e,1} = \begin{bmatrix} s_1 \\ H_{e,22}^\perp s_2 \end{bmatrix} \quad (7)$$

If $[H_{e,11}^T (H_{e,22}^\perp H_{e,21})^T]^T$ has full column rank, then $z_{e,1}$ can be recovered exactly. The decoding algorithm now suggests itself, i.e., find the smallest possible $H_{e,22}$ such that $[H_{e,11}^T (H_{e,22}^\perp H_{e,21})^T]^T$ has full rank and it is outlined in Algorithm 1.

Algorithm 1 Decoder for the BEC

1) Suppose, at time t , the earliest uncorrected error is at a delay d . Identify z_e and H_e as defined above.

2) Starting with $d' = 1, 2, \dots, d$, partition

$$z_e = [z_{e,1}^T, z_{e,2}^T]^T \text{ and } H_e = \begin{bmatrix} H_{e,11} & 0 \\ H_{e,21} & H_{e,22} \end{bmatrix}$$

where $z_{e,2}$ correspond to the erased positions up to delay d' .

3) Check whether the matrix $\begin{bmatrix} H_{e,11} \\ H_{e,22}^\perp H_{e,21} \end{bmatrix}$ has full column rank.

4) If so, solve for $z_{e,1}$ in the system of equations

$$\begin{bmatrix} H_{e,11} \\ H_{e,22}^\perp H_{e,21} \end{bmatrix} z_{e,1} = \begin{bmatrix} s_1 \\ H_{e,22}^\perp s_2 \end{bmatrix}$$

5) Increment $t = t + 1$ and continue.

A. Complexity

Suppose the earliest uncorrected error is at time $t - d + 1$, then steps 2), 3) and 4) in Algorithm 1 can be accomplished by just reducing H_e into the appropriate row echelon form, which has complexity $O(d^3)$. The earliest entry in z_e is at time $t - d + 1$ implies that it was not corrected at time $t - 1$, the probability of which is $P_{d-1,t-1}^e \leq \eta 2^{-n\beta(d-1)}$. Hence, the average decoding complexity is at most $K \sum_{d>0} d^3 2^{-n\beta d}$ which is bounded and is independent of t . In particular, the probability of the decoding complexity being Kd^3 is at most $\eta 2^{-n\beta d}$. The decoder is easy to implement and its performance is simulated in Section VIII. Note that the encoding complexity per time iteration increases linearly with time. This can also be made constant on average if the decoder can send periodic acks back to the encoder with the time index of the last correctly decoded source bit.

V. SUFFICIENT CONDITIONS FOR STABILIZABILITY

Consider an unstable m -dimensional linear system whose state space equations in canonical form are given by (1), i.e., $\lambda(F) > 1$, and recall that the characteristic polynomial of F is $z^n + a_1 z^{n-1} + \dots + a_m$. Suppose the observer does not have any feedback from the controller, in particular, it does not have access to the control inputs. Then we can stabilize such a system in the mean squared sense over a noisy channel provided that the rate R and exponent β of the (R, β) -anytime reliable code used to encode the measurements satisfy the following sufficient condition.

Theorem 5.1 (No Feedback to the Observer): It is possible to stabilize (1) in the mean squared sense with an (R, β) -anytime code provided (F, B) is controllable and

$$R > R_n = \frac{1}{n} \log_2 \sum_{i=1}^m |a_i|, \quad \beta > \beta_n = \frac{2}{n} \log_2 \lambda(\bar{F}) \quad (8)$$

If the observer knows the control inputs, it turns out that one can make do with lower rates. This is stated as the following Theorem

Theorem 5.2 (Observer Knows the Control Inputs): When the observer has access to the control inputs, it is possible to stabilize (1) in the mean squared sense with an (R, β) -anytime code provided (F, B) is controllable and

$$R > R_n^f = \argmin_r \{ \lambda(\bar{F} D_{nr}) < 1 \} \quad (9a)$$

$$\beta > \beta_n^f = \frac{2}{n} \log_2 \lambda(\bar{F}) \quad (9b)$$

where $D_{nr} = \text{diag}(2^{-nr}, 1, \dots, 1)$. Moreover

$$R_n^f \leq \frac{1}{n} \log_2 \max \left\{ |a_m| 2^{m-1}, \max_{1 \leq i \leq m-1} |a_i| 2^i \right\} \quad (10)$$

The superscript f in R_n^f denotes ‘feedback’ to emphasize the fact that the observer has access to the control inputs. Before proceeding further, we will give a brief outline of the proofs for Theorems 5.1 and 5.2 (details are in Section VII). At each time t , using the channel outputs received till t , we bound the set of all possible states that are consistent with the estimates of the quantized measurements using a hypercuboid, i.e., a

region of the form $\{x_t \in \mathbb{R}^m | x_{\min,t|t} \leq x_t \leq x_{\max,t|t}\}$, where $x_{\min,t|t}, x_{\max,t|t} \in \mathbb{R}^m$ and the inequalities are component-wise. If $\Delta_{t|t} = x_{\max,t|t} - x_{\min,t|t}$, then from Lemma 7.1, $\Delta_{t+1|t} = \bar{F}\Delta_{t|t} + W\mathbf{1}_m$. The anytime exponent is determined by the growth of Δ_t in the absence of measurements, hence the bound $\beta_n = \beta_n^f = 2\log_2 \lambda(\bar{F})$. The bound on the rate is determined by how fine the quantization needs to be for Δ_t to be bounded asymptotically.

A. The Limiting Case

The sufficient conditions derived above are for the case when the measurements are encoded every time step. Alternately, one can encode the measurements every, say ℓ , time steps, and consider the asymptotic rate and exponent needed as ℓ grows. Note that this amounts to working with the system matrix F^ℓ . So, one can calculate this limiting rate and exponent by writing the eigen values of F , $\{\lambda_i\}_{i=1}^m$, as $\lambda_i = \mu_i^n$ and letting n scale. The following asymptotic result allows us to compare the sufficient conditions above with those in the literature (eg., see [3], [5], [11]).

Theorem 5.3 (The Limiting Case): Write the eigen values of F , $\{\lambda_i\}_{i=1}^m$, in the form $\lambda_i = \mu_i^n$. Letting n scale, R_n and R_n^f converge to R^* , and β_n and β_n^f converge to β^* , where

$$R^* = \sum_{i:|\mu_i|>1} \log_2 |\mu_i|, \quad \beta^* = 2\log_2 \max_i |\mu_i| \quad (11)$$

In addition, the upper bounds on R_n^f in (10) also converges to R^* .

Proof: See Section C of the Appendix. ■

For stabilizing plants over deterministic rate limited channels, [3] showed that a rate $R > R^*$, where R^* is as in (11), is necessary and sufficient. So, asymptotically the sufficient conditions for the rate R in Theorems 5.1 and 5.2 are tight. Though the above limiting case allows one to obtain a tight and an intuitively pleasing characterization of the rate and exponent needed, it should be noted that this may not be operationally practical. For, if one encodes the measurements every ℓ time steps, even though Theorem 5.3 guarantees stability, the performance of the closed loop system (the LQR cost, say) may be unacceptably large because of the delay we incur. This is what motivated us to present the sufficient conditions in the form that we did above.

B. A Comment on the Trade-off Between Rate and Exponent

Once a set of rate-exponent pairs (R, β) that can stabilize a plant is available, one would want to identify the pair that optimizes a given cost function. Higher rates provide finer resolution of the measurements while larger exponents ensure that the controller's estimate of the plant does not drift away; however, we cannot have both. One can either coarsely quantize the measurements and protect the bits heavily or quantize them moderately finely and not protect the bits as much. One can easily cook up examples using an LQR cost function with the balance going either way. Studying this trade-off is integral to making the results practically applicable.

VI. TIGHTER BOUNDS ON THE ANYTIME EXPONENT

From Theorem 5.1, using the technique outlined in the previous section, one needs an exponent $n\beta \geq 2\log \lambda(\bar{F})$. It turns out that a smaller exponent of $2\log_2 \lambda(F)$ suffices. The idea is to alternately bound the set of all possible states that are consistent with the estimates of the quantized measurements using an ellipsoid $\mathcal{E}(P, c) \triangleq \{x \in \mathbb{R}^m | \langle x - c, P^{-1}(x - c) \rangle \leq 1\}$. This can be seen as an extension of the technique proposed in [12] to filtering using quantized measurements. If $m = 1$, $\lambda(\bar{F}) = \lambda(F)$. So, let $m \geq 2$.

In view of the duality between estimation and control, we can focus on the problem of tracking (1) over a noisy communication channel. For, if (1) can be tracked with an asymptotically finite mean squared error and if (F, B) is stabilizable, then it is a simple exercise to see that there exists a control law $\{u_t\}$ that will stabilize the plant in the mean squared sense, i.e., $\limsup_t \mathbb{E}\|x_t\|^2 < \infty$. In particular, if the control gain K is chosen such that $\sqrt{2}F + BK$ is stable, then $u_t = K\hat{x}_{t|t}$ will stabilize the plant, where $\hat{x}_{t|t}$ is the estimate of x_t using channel outputs up to time t . Hence, in the rest of the analysis, we will focus on tracking (1). The control input u_t therefore is assumed to be absent, i.e., $u_t = 0$.

We will first present a recursive state estimation algorithm using the channel outputs and then state the sufficient conditions needed for the estimation error to be appropriately bounded using such a filter. Recall that the channel outputs corresponding to the coded bits $c_t \in GF_2^n$ are $z_t \in \mathcal{Z}^n$. Let $x_0 \in \mathcal{E}(P_0, 0)$ and suppose using $\{z_\tau\}_{\tau \leq t-1}$, we have $x_t \in \mathcal{E}(P_{t|t-1}, \hat{x}_{t|t-1})$. Note that, since $H = [1, 0, \dots, 0]$, the measurement update provides information of the form $x_{\min,t|t}^{(1)} \leq x_t^{(1)} \leq x_{\max,t|t}^{(1)}$, which one may call a slab. $\mathcal{E}(P_{t|t}, \hat{x}_{t|t})$ would then be an ellipsoid that contains the intersection of the above slab with $\mathcal{E}(P_{t|t-1}, \hat{x}_{t|t-1})$, in particular one can set it to be the minimum volume ellipsoid covering this intersection. Lemma A.1 gives a formula for the minimum volume ellipsoid covering the intersection of an ellipsoid and a slab. Note that the width of the slab above tends to be smaller if the observer has access to the control inputs than when it does not. For the time update, it is easy to see that for any $\epsilon > 0$ and $P_{t+1} = (1+\epsilon)FP_{t|t}F^T + \frac{W^2}{4\epsilon}\mathbf{1}_m$, $\mathcal{E}(P_{t+1}, F\hat{x}_{t|t})$ contains the state x_{t+1} whenever $\mathcal{E}(P_{t|t}, \hat{x}_{t|t})$ contains x_t . This leads to the following Lemma. For convenience, we write P_t for $P_{t|t-1}$.

Lemma 6.1 (The Ellipsoidal Filter): Whenever $\mathcal{E}(P_0, 0)$ contains x_0 , for each $\epsilon > 0$, the following filtering equations give a sequence of ellipsoids $\{\mathcal{E}(P_{t|t}, \hat{x}_{t|t})\}$ that, at each time t , contain x_t .

$$P_{t+1} = (1+\epsilon)FP_{t|t}F^T + \frac{W^2}{4\epsilon}\mathbf{1}_m, \quad \hat{x}_{t+1} = F\hat{x}_{t|t} \quad (12a)$$

$$P_{t|t} = b_t P_t - (b_t - a_t) \frac{P_t e_1 e_1^T P_t}{e_1^T P_t e_1}, \quad \hat{x}_{t|t} = \xi_t \frac{P_t e_1}{\sqrt{e_1^T P_t e_1}} \quad (12b)$$

where a_t, b_t and ξ_t can be calculated in closed form using Lemma A.1.

Using this approach, we get the following set of sufficient conditions. The proofs are similar to the proofs of Theorems

5.1 and 5.2, and hence skipped due to space limitations.

Theorem 6.2 (No Feedback to the Observer): It is possible to stabilize (1) for $m \geq 2$ in the mean squared sense with an (R, β) -anytime code provided (F, B) is controllable and

$$R > R_{e,n} = \frac{1}{n} \log_2 \left[\frac{\sqrt{m}}{2} \sum_{i=1}^m |a_i| \theta^{i-1} \right] \quad (13a)$$

$$\beta > \beta_{e,n} = \frac{2}{n} \log_2 \lambda(F) \quad (13b)$$

where $\theta = \frac{m}{m-1}$

Theorem 6.3 (Observer Knows the Control Inputs): When the observer has access to the control inputs, it is possible to stabilize (1) in the mean squared sense with an (R, β) -anytime code provided (F, B) is controllable and

$$R > R_{e,n}^f = \operatorname{argmin}_r \{ \lambda(\bar{F} D_{m,nr}) < 1 \} \quad (14a)$$

$$\beta > \beta_{e,n}^f = \frac{2}{n} \log_2 \lambda(F) \quad (14b)$$

where $D_{m,nr} = \operatorname{diag}(\sqrt{m} 2^{-nr}, \sqrt{\theta}, \dots, \sqrt{\theta})$, $\theta = \frac{m}{m-1}$. Moreover

$$R_{e,n}^f \leq \frac{1}{2n} \log_2 m + \frac{1}{n} \log_2 \max \left\{ |a_m| (2\theta)^{m-1}, \max_{1 \leq i \leq m-1} 2|a_i| (2\theta)^{i-1} \right\} \quad (15)$$

In the same limiting sense as described in Section V, $R_{e,n}^f$ and $R_{e,n}$ converge to R^* while $\beta_{e,n}^f$ and $\beta_{e,n}$ converge to β^* , where R^* and β^* are as in the Lemma 5.3. The proof is in Section C of the Appendix.

VII. PROOFS OF THEOREMS 5.1 AND 5.2

The analysis will proceed in two steps. We will first determine a sufficient condition on the number of bits per measurement, nR , that are required to track (1) when these bits are available error free. We will then determine the anytime exponent $n\beta$ needed in decoding these source bits when they are communicated over a noisy channel.

At each time, we bound the set of all possible states that are consistent with the quantized measurements using a hypercuboid, i.e., a region of the form $\{x \in \mathbb{R}^m | x_{\min} \leq x \leq x_{\max}\}$, where $x_{\min}, x_{\max} \in \mathbb{R}^m$ and the inequalities are component-wise. In what follows, we call $\Delta_{t|\tau} \triangleq x_{\max,t|\tau} - x_{\min,t|\tau}$, the uncertainty in x_t using $\{b'_\tau\}_{\tau' \leq \tau}$, i.e., quantized measurements up to time τ . For convenience, let $\Delta_t \equiv \Delta_{t|t-1}$. Then, the time update is given by the following Lemma.

Lemma 7.1 (Time Update): The time update relating Δ_{t+1} and $\Delta_{t|t}$ is given by $\Delta_{t+1} = \bar{F} \Delta_{t|t} + W \mathbf{1}_m$

Proof: From the system dynamics in (1), the following is immediate

$$\begin{aligned} \Delta_{t+1}^{(i)} &= W + \max \left\{ \left| \pm a_i \Delta_{t|t}^{(1)} + \Delta_{t|t}^{(i+1)} \right|, \left| \Delta_{t|t}^{(i+1)} \right|, \left| a_i \Delta_{t|t}^{(1)} \right| \right\} \\ &= |a_i| \Delta_{t|t}^{(1)} + \Delta_{t|t}^{(i+1)} + W, \quad i \leq m-1 \\ \Delta_{t+1}^{(m)} &= |a_m| \Delta_{t|t}^{(1)} + W \end{aligned}$$

In short, the above equations amount to $\Delta_{t+1} = \bar{F} \Delta_{t|t} + W \mathbf{1}_m$. \blacksquare

The measurement update depends on whether or not the observer has access to the control inputs.

A. Observer does not know the control inputs

In this case, the observer simply quantizes the measurements y_t according to a 2^{nR} -regular lattice quantizer with bin width δ , i.e., the quantizer is defined by $Q: \mathbb{R} \mapsto \{0, 1, \dots, 2^{nR} - 1\}$, where $Q(x) = \lfloor \frac{x}{\delta} \rfloor \bmod 2^{nR}$. Assuming that the rate, R , is large enough, we will first find the steady state value of the recursion for Δ_t , which we then use to determine R . At each time t , the observer can communicate the measurement y_t to within an uncertainty of δ , i.e., the estimator knows that the measurement lies in an interval of width δ . Adding to this the effect of the observation noise, $-\frac{V}{2} \leq v_t \leq \frac{V}{2}$, the estimator knows $x_t^{(1)}$ to within an uncertainty of $\Delta_{t|t}^{(1)} = \delta + V$. Note that $\Delta_{t|t}^{(i)} = \Delta_t^{(i)}$ for $i \neq 1$. Combining this observation with Lemma 7.1, the following is fairly straightforward.

Lemma 7.2 (Steady State value of Δ_t without feedback): If $\lim_{t \rightarrow \infty} \Delta_t = \Delta_\infty$, then $\Delta_\infty = (\delta + V) L_u a + W L_u \mathbf{1}_m$, where $a = [a_1, \dots, a_m]^T$ and $L_u = [\ell_{ij}]_{1 \leq i, j \leq m}$ with $\ell_{ij} = \mathbb{I}_{i \leq j}$.

Now, we need to go back and calculate R . Observe that Δ_∞ does not depend on the starting value Δ_0 . So we just need $\delta 2^{nR} \geq \max \{ \Delta_\infty^{(1)}, \Delta_0^{(1)} \} + V$. From the above Lemma, $\Delta_\infty^{(1)} = \delta \sum_{i=1}^m |a_i| + V \sum_{i=1}^m |a_i| + mW$. So, we need

$$2^{nR} > \max \left\{ \sum_{i=1}^m |a_i| + \frac{V + V \sum_{i=1}^m |a_i| + mW}{\delta}, \frac{\Delta_0^{(1)}}{\delta} \right\}$$

The minimum required rate is obtained by letting $\delta \rightarrow \infty$, in which case we need $R > \frac{1}{n} \log_2 \sum_{i=1}^m |a_i|$ and this gives R_n in Theorem 5.1.

B. Observer knows the control inputs

In this case, the observer can infer that the uncertainty in y_t at the estimator side is $\Delta_t^{(1)} + V$. So, it can use the nR bits to shrink this to $2^{-nR}(\Delta_t^{(1)} + V)$. Taking into account the observation noise, the uncertainty in x_t after the measurement update will be given by $\Delta_{t|t}^{(1)} = V + 2^{-nR}(\Delta_t^{(1)} + V)$ and $\Delta_{t|t}^{(i)} = \Delta_t^{(i)}$ for $i \neq 1$. Combining this with Lemma 7.1, the overall recursion for Δ_t is given by

$$\begin{aligned} \Delta_{t+1} &= \bar{F} D_{nR} \Delta_t + W_{c,nR}, \quad \text{where} \\ D_{nR} &= \operatorname{diag}\{2^{-nR}, 1, \dots, 1\} \\ W_{c,nR} &= [V(1 + 2^{-nR}) + W, W, \dots, W]^T \end{aligned} \quad (16)$$

Noting that $V(1 + 2^{-nR}) \leq 2V$, the above recursion is bounded if and only if $\bar{F} D_{nR}$ is stable. It follows that $\bar{F} D_{nR}$ is stable for all $R > R_n^f$, where $R_n^f = \frac{1}{n} \operatorname{argmin}_r \{ \lambda(\bar{F} D_{nr}) < 1 \}$.

Now consider tracking (1) over a noisy channel. Intuitively, the desired anytime exponent is determined only by the growth of the tracking error in the absence of measurements, which by Lemma 7.1 is governed by \bar{F} . This is independent of whether or not control input is available at the observer. This explains the value of $\beta_n = \beta_n^f = \frac{2}{n} \log_2 \lambda(\bar{F})$ in Theorems 5.1 and

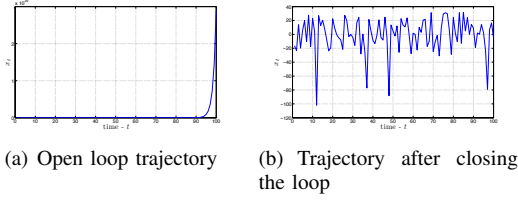


Fig. 2.

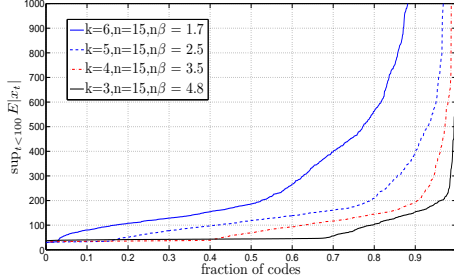


Fig. 3. The control performance of the code ensemble improves as the rate decreases

5.2. Making this argument rigorous is simple and has not been presented here due to space limitations.

VIII. SIMULATIONS

We present two examples, one scalar and one vector, and stabilize them over a binary erasure channel with erasure probability $\epsilon = 0.3$. The number of channel uses per measurement is fixed to $n = 15$. In both cases, time invariant codes $\mathbb{H}_{15,R} \in \mathbb{TZ}_{\frac{1}{2}}$, for an appropriate rate R , were randomly generated and decoded using Algorithm 1.

A. Example 1

Consider stabilizing the scalar unstable process obtained by setting $m = 1$, $-a_1 = 2$ in (1) with w_t and v_t being uniform on $[-30, 30]$ and $[-1, 1]$ respectively. Using Theorem 5.1, in order to stabilize x_t in the first moment sense, one needs a code with exponent $\beta \geq \frac{1}{n} = 0.0667$ and $k = nR \geq 1$. Using Theorem 3.2, causal linear codes exist for $\beta < \beta^* = H^{-1}(1 - R) \left(\log_2 \left(\frac{1}{\zeta} \right) + \log_2 (2^{1-R} - 1) \right)$. A quick calculation shows that for $k = 6, n = 15$, $\beta^* = \frac{1.1413}{n} = 0.0761 > 0.0667$. The observer does not have access to the control inputs, so an s^k -regular lattice quantizer with bin-width δ was used to quantize the measurements. The control input is just $u_t = -\hat{x}_{t|t-1}$. The four curves in Fig 3 correspond to the following sets of values: $(k = 3, \delta = 16)$, $(k = 4, \delta = 8)$, $(k = 5, \delta = 4)$ and $(k = 6, \delta = 2)$. Fig 2 shows the plot of a sample path of the above process with $k = 3, \delta = 16, L = 2^3$ before and after closing the loop, the fact that the plant has been stabilized is clear. By easing up on the rate R , i.e., by performing coarser quantization but better error correction, the control performance of the code ensemble improves. This is demonstrated in Fig 3. For each value of k from 3 to 6, 1000 time invariant codes were generated at random from $\mathbb{TZ}_{\frac{1}{2}}$. Each such code was used to control the process above over a time horizon of

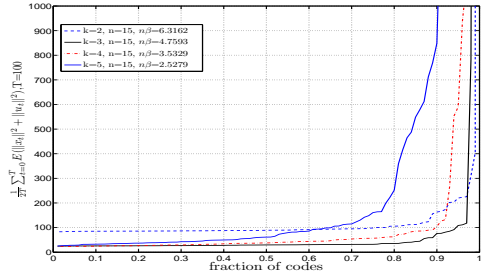


Fig. 4. The CDF of the LQR costs for different realizations of the codes

$T = 100$. The x -axis denotes the proportion of codes for which $\sup_{t < 100} \mathbb{E}|x_t|$ is below a prescribed value, e.g., with $k = 6, n = 15$, $\sup_{t < 100} \mathbb{E}|x_t|$ was less than 200 for 50% of the codes while with $k = 3, n = 15$, this fraction increases to more than 95%. The y -axis has been capped at 1000 for clarity. This shows that one can tradeoff utilization of communication resources and control performance.

B. Example 2

Consider a 3-dimensional unstable system (1) with $a_1 = -2$, $a_2 = -0.25$, $a_3 = 0.5$ and $B = \mathcal{I}_3$. Each component of w_t and v_t is generated i.i.d $N(0, 1)$ and truncated to $[-2.5, 2.5]$. The eigen values of F are $\{2, -0.5, 0.5\}$ while $\lambda(\bar{F}) = 2.215$. The observer has access to the control inputs and we use the hypercuboidal filter outlined in Section VII. Using Theorem 5.2, the minimum required bits and exponent are given by $k = nR \geq 2$ and $n\beta \geq 2 \log_2 2.215 = 2.29$. The control input is $u_t = -\hat{x}_{t|t-1}$. For $k \leq 5$, $n\beta \geq 2.53$. The competition between the rate and the exponent in determining the LQR cost is evident when we look at the LQR cost $\frac{1}{200} \sum_{i=1}^{100} \mathbb{E} [\|x_t\|^2 + \|u_t\|^2]$ in Fig 4. When $k = 2$, the error exponent $n\beta = 6.3$ is large. So, at any time t , the decoder decodes all the source bits $\{b_\tau\}_{\tau \leq t-1}$ with a high probability. Hence, the limiting factor on the LQR cost is the resolution the source bits b_t provide on the measurements. But when $k = 5$, the measurements are available almost losslessly but the decoder makes errors in decoding the source bits. Fig 4 suggest that the best choice of rate is $R = 3/15 = 0.2$.

IX. CONCLUSION

We presented an explicit construction of anytime reliable tree codes with efficient encoding and decoding over erasure channels. We also gave several sufficient conditions on the rate and reliability required of the tree code to guarantee stability, and argued that they are asymptotically tight. Although the work described here is a major step towards controlling plants over noisy channels, there are many issues to study and resolve. The tradeoff between rate and reliability (how finely to quantize the measurements vs. how much error protection to use) to optimize system performance (such as an LQR cost) remains to be studied, as well as how best to quantize and generate control signals. Furthermore, the problem of constructing efficiently decodable tree codes for other classes of channels, such as the BSC and the AWGNC, remains open.

REFERENCES

- [1] J. Baillieul and P.J. Antsaklis, "Control and communication challenges in networked real-time systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 9–28, Jan 2007.
- [2] L.J. Schulman, "Coding for interactive communication," *Information Theory, IEEE Transactions on*, vol. 42, no. 6, pp. 1745–1756, 1996.
- [3] GN Nair and RJ Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM Journal on Control and Optimization*, vol. 43, no. 2, pp. 413–436, 2005.
- [4] Alexey S. Matveev and Andrey V. Savkin, *Estimation and Control over Communication Networks (Control Engineering)*, Birkhauser, 2007.
- [5] Anant Sahai and Sanjoy Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link - part i: Scalar systems," *Information Theory, IEEE Transactions on*, vol. 52, no. 8, pp. 3369–3395, 2006.
- [6] R Ostrovsky, Y Rabani, and LJ Schulman, "Error-correcting codes for automatic control," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 2931–2941, 2009.
- [7] Serdar Yuksel, "A random time stochastic drift result and application to stochastic stabilization over noisy channels," *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 628–635, 2009.
- [8] T. Simsek, R. Jain, and P. Varaiya, "Scalar estimation and control with noisy binary observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1598–1603, 2004.
- [9] CE Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct 1948.
- [10] Ravi Teja Sukhvasi and Babak Hassibi, "Linear error correcting codes with anytime reliability," <http://arxiv.org/abs/1102.3526>, 2011.
- [11] P. Minero, M. Franceschetti, S. Dey, and G.N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *Automatic Control, IEEE Transactions on*, vol. 54, no. 2, 2009.
- [12] F. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *Automatic Control, IEEE Transactions on*, vol. 13, no. 1, Feb. 1968.
- [13] Osman Güler and Filiz Gürtuna, "The extremal volume ellipsoids of convex bodies, their symmetry properties, and their determination in some special cases," *arXiv*, vol. math.MG, Sep 2007.
- [14] A Sluis, "Upperbounds for roots of polynomials," *Numerische Mathematik*, vol. 15, no. 3, pp. 250–262, 1970.

APPENDIX

A. The Minimum Volume Ellipsoid

Lemma A.1 (Theorem 6.1 [13]): The minimum volume ellipsoid $\mathcal{E}(\hat{P}, c)$ covering

$$\{x \in \mathbb{R}^m | x \in \mathcal{E}(P, 0), \gamma \sqrt{h^T P h} \leq \langle h, x \rangle \leq \delta \sqrt{h^T P h}\}$$

where $|\delta| \geq |\gamma|$, is given by

$$\hat{P} = bP - (b-a) \frac{Phh^T P}{h^T P h}, \quad c = \xi \frac{Ph}{\sqrt{h^T P h}} \quad (17)$$

where

- 1) If $\gamma\delta < -\frac{1}{m}$, then $\xi = 0, a = b = 1$
- 2) If $\gamma + \delta = 0$ and $\gamma\delta > -\frac{1}{m}$, then

$$\xi = 0, \quad a = m\delta^2, \quad b = \frac{m(1-\delta^2)}{m-1}$$

- 3) If $\gamma + \delta \neq 0$ and $\gamma\delta > -\frac{1}{m}$, then

$$\xi = \frac{m(\gamma + \delta)^2 + 2(1 + \gamma\delta) - \sqrt{D}}{2(m+1)(\gamma + \delta)}$$

$$a = m(\xi - \gamma)(\xi - \delta), \quad b = \frac{a - a\gamma^2}{a - (\xi - \gamma)^2}$$

$$\text{where } D = m^2(\delta^2 - \gamma^2)^2 + 4(1 - \gamma^2)(1 - \delta^2)$$

If $|\delta| < |\gamma|$, change x to $-x$ and apply the above result.

B. Upper bounds on R_n^f and $R_{e,n}^f$

There are several bounds in the Mathematics literature on the roots of a polynomial in terms of the polynomial coefficients, a standard and near optimal bound being the Fujiwara's bound which we state below.

Lemma A.2 (Fujiwara's Bound): Consider the monic polynomial with complex coefficients $f(x) = x^m + c_1 x^{m-1} + \dots + c_m$ and let $\lambda(f)$ denote the largest root in magnitude. Then

$$\lambda(f) \leq K(f) = 2 \max \left\{ |c_1|, |c_2|^{\frac{1}{2}}, \dots, |c_{m-1}|^{\frac{1}{m-1}}, \left| \frac{c_m}{2} \right|^{\frac{1}{m}} \right\}$$

The upper bounds on R_n^f and $R_{e,n}^f$ can now be proved as follows. The characteristic polynomial of $\bar{F}D_{nr}$ is given by $f_{c,nr}(x) = x^m - 2^{-nr} \sum_{i=1}^m |a_i| x^{m-i}$. Applying Lemma A.2, if the rate R is larger than the smallest value of r that will make $K(f_{c,nr}) < 1$, then $\lambda(\bar{F}D_{nR}) \leq K(f_{c,nR}) < 1$ making $\bar{F}D_{nR}$ stable. The bound for R_n^f is then immediate while the bound for $R_{e,n}^f$ follows by noting that the characteristic polynomial of $\bar{F}D_{m,nr}$ is $x^m - \sqrt{m} 2^{-nr} \sum_{i=1}^m \theta^{i-1} |a_i| x^{m-i}$.

C. The Limiting Case

Let F is any m -dimensional square matrix and $f(x)$ denotes its characteristic polynomial. Then the following bounds hold (for details see [14])

$$\lambda(F) \leq \lambda(\bar{F}) \leq \frac{\lambda(F)}{\sqrt[m]{2} - 1}, \quad K(f) \leq 2\lambda(\bar{F}) \quad (18)$$

The proof for $\lim_{n \rightarrow \infty} \beta_n^f = \beta^*$ follows easily from the first bound in (18).

By the hypothesis of the Lemma, the eigen values of F_n are of the form $\{\mu_i^n\}_{i=1}^m$. To emphasize the fact that F depends on n , we write it as F_n and a_i as $a_{i,n}$. Recall that the characteristic polynomial of F_n is given by $f_n(x) = x^m + a_{1,n} x^{m-1} + \dots + a_{m,n}$. Let $\mathcal{I}_u \triangleq \{i \mid |\mu_i| \geq 1\}$, then the following is easy to prove

$$\lim_{n \rightarrow \infty} \frac{|a_{i,n}|}{|a_{|\mathcal{I}_u|,n}|} = 0, \quad i \neq |\mathcal{I}_u|, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 |a_{|\mathcal{I}_u|,n}| = \sum_{i \in \mathcal{I}_u} \log_2 |\mu_i| \quad (19)$$

We will prove that R_n and R_n^f converge to R^* , the proof for $R_{e,n}$ and $R_{e,n}^f$ is similar. From (19), it is obvious that $\lim_{n \rightarrow \infty} R_n = \sum_{i \in \mathcal{I}_u} \log_2 |\mu_i|$. It remains to show that the limit holds for R_n^f . The characteristic polynomial of $\bar{F}D_{nr}$ is given by $f_{c,nr}(x) = x^m - 2^{-nr} \sum_{i=1}^m |a_i| x^{m-i}$. From (C), we have $\frac{1}{2} K(f_{c,nr}) \leq \lambda(\bar{F}D_{nr}) \leq K(f_{c,nr})$. Define $R_{n,1}^f \triangleq \argmin_r \{ \frac{1}{2} K(f_{c,nr}) \leq 1 \}$ and $R_{n,2}^f \triangleq \argmin_r \{ K(f_{c,nr}) \leq 1 \}$. Then, it is obvious that $R_{n,1}^f \leq R_n^f \leq R_{n,2}^f$. Using Lemma A.2, some simple algebra and taking limit $n \rightarrow \infty$, we get

$$\lim_{n \rightarrow \infty} R_n^f = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \max \left\{ \frac{|a_{m,n}|}{2}, \max_{1 \leq i \leq m-1} |a_{i,n}| \right\}$$

Combining this with (19), we get the desired result, i.e., $\lim_{n \rightarrow \infty} R_n^f = \sum_{i \in \mathcal{I}_u} \log_2 |\mu_i|$.